

## Tilburg University

### The construction and use of approximations for missing quarterly observations

Nijman, T.E.; Palm, F.C.

*Published in:*

Journal of Business and Economic Statistics

*Publication date:*

1985

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Nijman, T. E., & Palm, F. C. (1985). The construction and use of approximations for missing quarterly observations: A model-based approach. *Journal of Business and Economic Statistics*, 4(1), 47-58.

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



# The Construction and Use of Approximations for Missing Quarterly Observations: A Model-Based Approach

**T. E. Nijman**

Tilburg University, Department of Econometrics, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

**F. C. Palm**

University of Limburg, Department of Economics, P.O. Box 616, 6200 MD Maastricht, The Netherlands.

We use a model-based approach to derive quarterly figures on several variables for the aggregate labor market in the Netherlands that are only observed annually. These approximations are conditional expectations derived from univariate and bivariate quarterly time series models for the series under consideration. They are subsequently used as proxies to estimate and analyze the structural labor market equations. Attention is given to the properties of estimation procedures based on proxy variables.

KEY WORDS: Proxy variables; Interpolation; Smoothing; Structural model; Kalman filter.

## 1. INTRODUCTION

Frequently in empirical economic modeling, observations are not available for the time periods for which the model is specified. For instance, for the aggregate labor market in the Netherlands, labor supply and employment are annually observed, whereas unemployment figures are available on a monthly basis. In the recently published quarterly macro-economic model for the Dutch economy [Central Planning Bureau (CPB) 1983], quarterly figures for the missing labor market data were derived by interpolation and subsequently used to estimate and analyze the quarterly model.

In this article we propose a model-based approach to the problem of missing observations and apply it to the aggregate labor market data for the Netherlands. Given a joint model for the data and for the missing observations, we show how to estimate the latter as mathematical expectations of the missing variable, conditional on the sample information. Usually, the model can be written in different forms, the role of which in deriving estimates of the missing observations will be discussed and illustrated. The derivation of approximations for the missing variables and their use in model building are closely interrelated. On the one hand, interpolation by means of conditional expectations requires parameter estimates. On the other hand, the parameters of the model can often be consistently estimated using the interpolated figures as proxies for the unobserved values, provided one takes into account the stochastic properties of the interpolations. This means, of course, that some information about the structure of the model is available. We shall explicitly pay attention to the relationship between interpolating the data and estimating the parameters in the model. In

particular, we consider consistent parameter estimation based on proxies for the unobserved variables. These methods are in general computationally attractive and fairly accurate. The terms *interpolations*, *approximations*, *proxy variables*, and *estimates of the missing observations* will be used as synonyms throughout this article.

A model-based approach has the advantage that the information available on the model and the series is incorporated in a coherent way in the analysis. Ideally, one would like to specify and estimate the joint model for the data and the missing observations and then derive interpolations for the missing values from the estimated model. In practice, however, the model is often not completely specified, or it is too large to be used to generate estimates of the missing figures. A limited information approach then has to be adopted in which the estimates of the missing observations and of the parameters are derived from part of the model only. Some forms of a structural model are adequate for the interpolation of the missing observations under limited information. The accuracy of the approximations generally depends on the amount of reliable information incorporated in the interpolation procedure. Other types of unobservables, such as measurement errors and expectational variables, can be approached along the same lines.

The article is organized as follows. Section 2 is devoted to methods for the interpolation of the missing observations. We also briefly review the properties of estimation procedures based on the interpolated values. We show how consistency of the estimates can be achieved. In Section 3, we present the structural model for the labor market. Sections 4 and 5 contain the empirical results. In Section 4.1, we specify quarterly univariate autoregressive integrated mov-



ing average (ARIMA) models for employment in the private and public sectors. These models are estimated from annual data and are used to generate smoothed estimates of the missing values. The use of indicator variables, which are observed quarterly, is considered in Section 4.2. The joint interpolation of several series is discussed in Section 4.3. Quarterly observations on unemployment are included to improve the smoothed estimates of the unobserved employment and labor supply figures. In Section 5, the newly constructed approximations for the employment and labor supply series are used as proxies to estimate a structural adjustment equation for labor demand. The results are compared with those obtained using the series constructed by the CPB as a proxy. The robustness of the empirical results is checked by allowing for measurement errors in the desired labor demand series. Section 6 presents some concluding remarks. Consistent estimators for the standard errors of the proxy variables estimators are given in the Appendix.

## 2. INTERPOLATION AND PARAMETER ESTIMATION

Consider the problem of interpolating the unobserved quarterly figures of a variable that is observed annually. A first approach consists in assuming that the quarterly values are generated by some univariate ARIMA model. This model could be interpreted as the marginal process implied by some linear quarterly simultaneous-equation model (SEM) for which the exogenous variables are generated by an ARIMA model (see Zellner and Palm 1974). By integrating the quarterly ARIMA model with respect to the missing observations, we get an ARIMA process for the annual data. The ARIMA model for the annual observations can be estimated from the data. The parameters of the quarterly ARIMA model can be estimated from the annual data, provided they are identified (see Palm and Nijman 1984). The conditional expectation of the quarterly values, given past annual observations or given all annual observations, can then be computed by means of the Kalman filter or the Wiener-Kolmogorov filtering theory to yield proxies for the missing variable. This approach has been adopted by Harvey and Pierse (1984) among others. The interpolation method put forward by Boot, Feibes, and Lisman (1967), which consists in minimizing the first or second squared difference of the interpolations given the sample information, is an approximation to this approach. It is exact when the appropriate model is a first- or second-order integrated scheme.

When a nonlinear SEM is appropriate, a univariate ARIMA model corresponds to or approximates the Wold representation of the quarterly or annual data. The normality assumption for the single variable may not be adequate in this case. Likelihood estimation methods based on the normal distribution then have to be interpreted as pseudolikelihood methods, and expectations are least squares approximations.

Second, related variables that are observed quarterly can be included in the conditioning set of the expectations of the missing variables. The regression model analyzed by Palm

and Nijman (1984) will then often be an appropriate tool for estimating the missing observations. It is likely that when the parameters of the conditional expectation of a quarterly variable, given its past annual values or past and future annual observations, are not identified, those of its conditional expectation given some indicator variable are identified. An indicator variable will then be required for interpolating the missing figures. If the prediction errors from conditioning on indicator variables only have an ARIMA correlation structure and are independent of these indicators, the errors can be interpolated along the lines of the first approach to get estimates of the missing observations that are conditional expectations given the sample information on the indicator variables and the variable to be interpolated. This procedure is a direct generalization of the procedures put forward by Chow and Lin (1971), Ginsburgh (1973), Fernandez (1981), and Litterman (1983).

Third, if several variables are not observed for all periods in the sample or if, for instance, their aggregate value is observed for some periods, one can model them jointly using a multivariate time series model or a structural SEM. In this way it becomes possible to take account of dependencies between the series and to derive sets of approximations for the missing figures that are internally consistent and that satisfy, for example, some adding-up restrictions across variables.

The amount of information incorporated in the interpolated series increases as we move from the first to the second and third approaches. In Section 4, the three approaches will be applied to aggregate labor market data for the Netherlands.

Now we briefly consider the estimation of the parameters in the model. When the joint density function for the data and the missing observations is given, the maximum likelihood (ML) method can be used. There are various ways to implement ML. By marginalizing with respect to the unobserved variables, one obtains the likelihood function, which can be maximized with respect to the parameters of the joint model, provided these parameters are identified. Some ML procedures do not require explicit marginalization with respect to the unobserved variables. For instance, the log-likelihood function in prediction error decomposition form and its derivatives can be evaluated by means of the Kalman filter to obtain ML estimates of the parameters and predictions for the missing data points. Similarly, the use of the EM algorithm put forward by Dempster, Laird, and Rubin (1977) avoids explicit marginalization of the model with respect to the unobserved variables.

Frequently, however, in applied work the model is too large to be jointly estimated or it is only specified in part. It may then be necessary to use a consistent estimation method instead of a fully efficient estimator. Consistent estimation procedures that are based on proxies for the missing figures are usually computationally attractive. [For details, we refer to Nijman and Palm (1984) and Pagan (1984).] When the proxies take into account important features of the model, their accuracy can be close to that of ML (see Nijman and



Palm 1984). Proxy variable estimators will often be more robust than the ML estimator, as they do not require a completely specified model.

To show how consistent proxy variable estimators can be obtained, consider the following regression model:

$$y_t = x_t' \beta + \varepsilon_t, \quad (2.1)$$

where  $y_t$  is the endogenous variable,  $x_t$  is a vector of exogenous variables,  $\beta$  is a vector of regression coefficients,  $\varepsilon_t$  is normally independently distributed with mean 0 and variance  $\sigma^2$ , and  $\varepsilon_t$  and  $x_{t'}$  are independent for all  $t, t'$ . Define  $x_{*t} = (y_t, x_t')'$ . Assume that the conditional expectation of  $x_{*t}$  given some information set  $z_t$  exists, and define

$$\bar{x}_{*t} = E(x_{*t}|z_t, \theta) \quad \text{and} \quad \hat{x}_{*t} = E(x_{*t}|z_t, \hat{\theta}), \quad (2.2)$$

where  $\theta$  and  $\hat{\theta}$  are vectors of parameters in the conditional expectation and their estimate, respectively.

When  $\bar{x}_{*t}$  is not observed for some or all  $t$ ,  $\hat{x}_{*t}$  can be used as a proxy for  $x_{*t}$ , provided a consistent estimate  $\hat{\theta}$  is available. The proxy equals the observed value, whenever the latter is available. Substitution of  $\hat{x}_{*t}$  into (2.1) yields

$$\hat{y}_t = \hat{x}_t' \beta + u_t, \quad (2.3)$$

where  $\hat{y}_t$  and  $\hat{x}_t$  are the appropriate elements of  $\hat{x}_{*t}$ ,  $u_t = \varepsilon_t + (x_{*t} - \hat{x}_{*t})' \beta$ , and  $\beta = (-1, \beta')'$ .

Ordinary least squares (OLS) applied to (2.3) will be consistent for  $\beta$  provided  $(\hat{X}'\hat{X})^{-1}\hat{X}'u$  converges to 0 in probability. Notice that mechanical interpolations that are not based on structural considerations, such as the method of Boot et al. (1967), usually do not satisfy this requirement. Palm and Nijman (1984) showed that the asymptotic bias can be important. Variables that are independent of  $\varepsilon_t$ , however, are also asymptotically independent of  $u_t$ , provided they are included in the conditioning set of (2.2). In this case, they are appropriate instrumental variables to estimate  $\beta$  consistently from (2.3). The larger the conditioning set is, the larger the number of valid instruments will be. Nijman and Palm (1984) showed how the efficiency of proxy variable estimators can be increased by incorporating additional information in the proxy. Finally, even if a regression of  $\hat{y}_t$  on  $\hat{x}_t$  yields a consistent estimate of  $\beta$ , the commonly used formula for the computation of standard errors of regression estimates will usually not yield consistent estimates of the standard errors. Consistent estimation of the standard errors of proxy variable estimators is discussed in the Appendix.

### 3. THE LABOR MARKET MODEL

The model for the labor market that we consider is a modified version of the quarterly CPB (1983) model. It describes the short-run adjustment of the labor demand and supply to desired labor demand and the trend in labor supply, respectively. Desired labor demand and the trend in labor supply depend on economic variables such as the real wage rate. In the present short-run model, however, they are assumed to be predetermined. Aggregate employment is determined by actual labor demand and supply. The model consists of three structural relationships, four definitions,

and an autoregressive integrated model for an exogenous variable. It reads as follows:

$$\begin{aligned} \gamma_1(L)\Delta d_t &= \delta_1(L)\Delta d_t^* \\ &+ \beta_1(d - d^*)_{t-1} + \varepsilon_{1t} \end{aligned} \quad (3.1)$$

$$\begin{aligned} \gamma_2(L)\Delta s_t &= \delta_2(L)\Delta s_t^* \\ &+ \beta_2(s - s^* - c_2)_{t-1} + \varepsilon_{2t} \end{aligned} \quad (3.2)$$

$$l_t = a_{1t}sb_t + a_{2t}d_t \quad (3.3)$$

$$n_t = l_t + g_t \quad (3.4)$$

$$sb_t = s_t - g_t \quad (3.5)$$

$$u_t = s_t - n_t \quad (3.6)$$

$$\gamma_3(L)\Delta g_t = \varepsilon_{3t}, \quad (3.7)$$

where  $d_t, d_t^*$  = labor demand and desired labor demand by the private sector, respectively;  $s_t, s_t^*$  = total labor supply and its trend value, respectively;  $l_t$  = employment in the private sector;  $sb_t$  = labor supply available for the private sector;  $g_t$  = employment in the public sector;  $n_t$  = total employment;  $u_t$  = unemployment;  $\gamma_i(L), \delta_i(L)$  denote finite-degree polynomials in the lag operator  $L$ , for which at most a second degree for  $\gamma_i(L)$  and a zeroth degree for  $\delta_i(L)$  will be sufficient in the sequel [i.e.,  $\gamma_i(L) = 1 - \gamma_{i1}L - \gamma_{i2}L^2$  and  $\delta_i(L) = \delta_i$ ];  $\Delta$  = the first difference operator;  $a_{it}$  = a variable weight that is assumed to be predetermined; and  $\varepsilon_{it}$  = a normally distributed white noise with mean 0 and constant variance  $\sigma_i^2$ . The  $\varepsilon_{it}$ 's are mutually independent.

A short explanation is in order. In (3.1) and (3.2), labor demand and supply are assumed to adjust to their respective target values  $d_t^*$  and  $s_t^* + c_2$  according to an error correction model (ECM). In the case of labor supply, this target value deviates from the trend value  $s_t^*$  because of a different treatment of some minor groups in the labor force in the definition of  $s_t$  and  $s_t^*$ . Notice that an adjustment according to (3.1) or (3.2) could be inappropriate when the target is nonstationary. However, the additional restriction  $\delta_i + \gamma_{i1} + \gamma_{i2} = 1$ , implying a zero mean lag, solves this problem for a linearly trending target (see Salmon 1982) and for a target variable generated by an ARIMA( $p, 1, q$ ) process (see Klok 1984).

Equation (3.3) is a linearized version of a relationship derived by Kooiman and Klok (1979), who aggregated over a normally distributed continuum of demand- or supply-constrained micro-labor markets to establish a nonlinear relationship between aggregate employment and aggregate excess labor supply. The weights  $a_{it}$  depend on excess labor supply in period  $t - 1$ . (For details, we refer to Kooiman and Klok 1979.) The main difference between the system (3.1)–(3.7) and the model of the CPB (1983) is that there  $d_t$  is assumed to be equal to  $d_t^*$ .

We analyze the quarterly model (3.1)–(3.7) for the period 1968–1981. In the empirical analysis we use the quarterly observations for  $u_t$ . The variables  $d_t^*, s_t^*, a_{1t}$ , and  $a_{2t}$  are not observed. It is reasonable to assume that these series are



smooth. The values for  $d_t^*$ ,  $a_{1t}$ , and  $a_{2t}$  generated in the model of the CPB are used as approximations. The variable  $d_t^*$  is determined by the real wage rate, the depreciation rate of capital, the rate of technical change, and new investments. The weights  $a_{1t}$  and  $a_{2t}$  depend on labor supply and demand in the private sector. The variable  $s_t^*$  constructed by the CPB is used. The determinants of  $s_t^*$  are labor supply, participation rates, and demographic variables. The labor demand ( $d_t$ ) is not observed and will be estimated in Section 4. For  $n_t$ ,  $g_t$ , and  $s_t$ , the annual average is observed. An overbar denotes the annual average [e.g.,  $\bar{n}_t = .25(n_t + n_{t-1} + n_{t-2} + n_{t-3})$ ]. Quarterly values of  $n_t$ ,  $g_t$ , and  $s_t$  will be estimated in the next section. The variables are measured in millions of man-years and have been seasonally adjusted. The data have been prepared by the CPB (1983) and are available on request. In Section 5.1, the errors involved in the approximations for  $d_t^*$ ,  $s_t^*$ ,  $a_{1t}$ , and  $a_{2t}$  are assumed to be negligible. In Section 5.2, this assumption will be weakened as far as  $d_t^*$  is concerned by adopting a measurement-errors framework.

#### 4. ESTIMATION OF THE MISSING OBSERVATIONS

##### 4.1 Estimates Derived From Univariate ARIMA Models

We start the empirical part of the analysis by interpolating total employment  $n_t$  for the Netherlands, assuming that the variable is generated by a univariate ARIMA scheme. This assumption is approximately in agreement with the structural model that we proposed in the previous section.

A first-order autoregressive (AR) model for the annual change in  $\bar{n}_t$  is found to be consistent with the information in the data (standard errors are given in parentheses):

$$\Delta_4 \bar{n}_t = .005 + .50 \Delta_4 \bar{n}_{t-4} + \hat{v}_t, \quad (.011) \quad (.30)$$

$$\hat{\sigma}_v^2 = 1.4 \times 10^{-3}, \quad (4.1)$$

for  $t \in T_4$ . By  $T_m$  we denote the set  $\{m, 2m, \dots, T\}$ , assuming for simplicity that  $T$  is a multiple of  $m$ .  $T_m^c$  denotes the set  $T_1 \setminus T_m$ . The disturbances of the various quarterly models will be denoted by  $\varepsilon_t$  if no confusion is possible, whereas  $v_t$  denotes the disturbance of the annual models. Because the sample size  $T/4 = 15$  is small, the information in the annual observations is probably limited.

Several models for  $n_t$ ,  $t \in T_1$ , are in agreement with (4.1). For instance, the specification

$$\Delta n_t = c_n + \rho \Delta n_{t-1} + \varepsilon_t, \quad t \in T_1, \quad (4.2)$$

implies the following model for  $\bar{n}_t$ ,  $t \in T_4$ :

$$\begin{aligned} (1 - \rho^4 L^4)(1 - L^4)\bar{n}_t \\ = .25(1 + \rho L + \rho^2 L^2 + \rho^3 L^3) \\ \times (1 + L + L^2 + L^3)^2(\varepsilon_t + c_n), \end{aligned} \quad (4.3)$$

which is identical to (4.1) except for the presence of a second-order moving average (MA) polynomial in  $L^4$  in (4.3).

Notice that it may be difficult to detect small MA coefficients such as those in (4.3) from the 15 annual observations.

ML estimation of (4.2) from the annual data using the prediction error decomposition method (e.g., see Harvey 1981) yields

$$\begin{aligned} (1 - .80L)(\Delta n_t - .000) &= \hat{\varepsilon}_t, \\ (.09) \quad (.005) \end{aligned} \quad \hat{\sigma}_\varepsilon^2 = 6.0 \times 10^{-5}, \quad (4.4)$$

with a log-likelihood value of 32.4. The numerical optimization for ML estimation was initialized at  $\hat{\rho} = (.50)^{.25} = .84$ , obtained from (4.1). A diffuse density function was assumed for the presample values of  $n_t$ , as proposed by Ansley and Kohn (1983). The standard errors have been computed using an expression for the information matrix given in Watson and Engle (1983), which requires only first derivatives of the log-likelihood function. Details on the computational aspects of ML estimation are given in Nijman (1985).

Alternatively, the model

$$\begin{aligned} (1 - .34L^4)(\Delta n_t - .003) &= \hat{\varepsilon}_t, \\ (.29) \quad (.004) \end{aligned} \quad \hat{\sigma}_\varepsilon^2 = 4.0 \times 10^{-4}, \quad t \in T_1, \quad (4.5)$$

with a log-likelihood value of 31.4 is in agreement with (4.1), since it implies an ARMA(1, 1) model in the lag  $L^4$  for  $\Delta_4 \bar{n}_t$  with a small MA coefficient. We are confronted with an identification problem. Several ARMA models for the quarterly data are approximately consistent with (4.1) for annual data. The choice between these models has to be based on prior information. We choose the model (4.4) because it is consistent with the structural model set out in Section 3 and because the data are seasonally adjusted. This model is used to compute two-sided conditional expectations for  $n_t$ ,  $t \in T_1$ , given  $\bar{n}_t$ ,  $t \in T_4$ , by means of the fixed interval smoother (see Anderson and Moore 1979). The approximations will be presented in Figure 1.

Although limited, the information in the annual data is not negligible. The random walk model, which implicitly underlies the interpolation scheme proposed by Boot et al. (1967), when  $d = 1$ , is rejected at the margin against the model (4.4) because it has a log-likelihood value of 30.1. When a Box-Pierce test is applied to the estimated innovations of the random walk model, no misspecification is detected. This test, however, is known to be conservative.

An explanation of the information in the data might be provided by the presence of the factor  $(1 + L + L^2 + L^3)^2$  in (4.3). As shown by Palm and Nijman (1984), the asymptotic efficiency of the ML estimate of  $\rho$  for complete data compared to that for an incomplete sample is 2.3 when  $\rho = .8$ . Therefore the information in the 14 annual changes equals that of  $2.3 \times 14 = 32$  observed quarterly changes. The presence of the factor  $(1 + L + L^2 + L^3)^2$  also explains why multiple maxima of the log-likelihood function were not found here.

A quarterly AR(1) model has also been assumed for the annual changes in the employment in the public sector  $g_t$ .



The estimated annual model is

$$\Delta_4 \bar{g}_t = .15 \Delta_4 \bar{g}_{t-4} + .011 + \hat{v}_t, \quad (27) \quad (.004)$$

$$\hat{\sigma}_v^2 = 1.7 \times 10^{-5}, \quad (4.6)$$

where  $\bar{g}_t$  is the annual average of  $g_t$ . The result in (4.6) approximately corresponds to the first-order AR model (3.7) for  $g_t$ ,  $t \in T_1$ , with  $\hat{\gamma}_{31} = .62$ . Estimating the model for  $\Delta g_t$  by ML from annual data yields

$$(1 - .57L)(\Delta g_t - .003) = \hat{\varepsilon}_t, \quad (12) \quad (.000)$$

$$\hat{\sigma}_\varepsilon^2 = 1.5 \times 10^{-6}. \quad (4.7)$$

The model (4.7) has been used to approximate the unobserved quarterly values of  $g_t$  by two-sided expectations conditional on the observed  $\bar{g}_t$ ,  $t \in T_4$ .

## 4.2 Interpolation by Means of Indicator Variables

Since one of the main determinants of the labor supply  $s_t$  is the trend  $s_t^*$ , which we assumed to be observed quarterly, the dynamic regression equation (3.2) can be used to construct approximations for the unobserved values of  $s_t$ . ML estimation of (3.2) yields

$$\begin{aligned} \Delta s_t &= .912 \Delta s_{t-1} - .024[s - s^* - .130]_{t-1} \\ &+ .106 \Delta s_t^* + \hat{\varepsilon}_{2t}, \end{aligned} \quad (18) \quad (.003) \quad (.007) \quad (20)$$

$$\hat{\sigma}_2^2 = 4.1 \times 10^{-7}, \quad (4.8)$$

with a log-likelihood value of 57.80.

The mean lag of (4.8) equals  $(\gamma_{21} + \delta_2 - 1)\beta_2^{-1} = -.75$ . Stationarity of error-correction terms in the presence of a trending target requires a zero mean lag. If we impose the constraint  $\gamma_{21} + \delta_2 = 1$  on (4.8), we find that

$$\begin{aligned} \Delta s_t &= .909 \Delta s_{t-1} - .025[s - s^* - .134]_{t-1} \\ &+ (1 - .909) \Delta s_t^* + \hat{\varepsilon}_{2t}, \end{aligned} \quad (17) \quad (.003) \quad (.004) \quad (17)$$

$$\hat{\sigma}_2^2 = 4.2 \times 10^{-7}, \quad (4.9)$$

with a log-likelihood value of 57.77. Evidently this model is close to the unrestricted version (4.8). This is not surprising if we realize that the value of the mean lag in (4.8) is strongly influenced by lags in a distant past.

## 4.3 Joint Interpolation of Several Series

Equation (4.9) can be used to compute the expectations of the unobserved quarterly values of  $s_t$  conditional on the annual observations on  $s_t$  and the quarterly observations on  $s_t^*$ . These approximations for  $s_t$  and the approximations for  $n_t$  obtained from (4.2) are not consistent with the observed quarterly data on unemployment because they do not satisfy identity (3.6). This problem can be solved by specifying a

joint model for labor supply and employment and adding the quarterly unemployment data to the information set used to compute approximations. If the changes in employment  $n_t$  and labor supply  $s_t$  are approximately independent, the unemployment series  $u_t$  is informative about these variables.

First, we determine a univariate ARIMA model for the seasonally adjusted unemployment series  $u_t$ . Models for  $n_t$  and  $s_t$  that are not in agreement with that for  $u_t$  can be safely ignored. The following model is found to be appropriate:

$$\Delta u_t = .0005 + 1.65 \Delta u_{t-1} - .73 \Delta u_{t-2} + \hat{\varepsilon}_t - .43 \hat{\varepsilon}_{t-1}, \quad (.0004) \quad (.24) \quad (.24) \quad (.30)$$

$$\hat{\sigma}_\varepsilon^2 = 2.1 \times 10^{-5}. \quad (4.10)$$

The AR polynomial in (4.10) has roots equal to  $(.82 \pm .22i)$ . Independent AR(1) models for  $\Delta n_t$  and  $\Delta s_t$  imply an ARMA(2, 1) model with real roots for  $\Delta u_t$ :

$$\begin{aligned} (1 - \rho_n L)(1 - \rho_s L) \Delta u_t \\ = c_u + (1 - \rho_n L) \varepsilon_{st} - (1 - \rho_s L) \varepsilon_{nt}, \end{aligned} \quad (4.11)$$

where the subscripts refer to the univariate processes for  $n_t$  and  $s_t$ , respectively. If we ignore the imaginary part of the roots, the result in (4.10) is in agreement with AR(1) processes for  $\Delta n_t$  and  $\Delta s_t$ , with  $\rho_n \approx \rho_s$ . Alternatively, AR(2) models for  $\Delta n_t$  and  $\Delta s_t$  with approximately the same AR parts imply a specification close to (4.10) for  $\Delta u_t$ .

We estimate the bivariate model

$$(1 - \rho_{n1} L - \rho_{n2} L^2)(\Delta n_t - c_n) = \varepsilon_{nt} \quad (4.12)$$

and

$$(1 - \rho_{s1} L - \rho_{s2} L^2)(\Delta s_t - c_s) = \varepsilon_{st}, \quad (4.13)$$

where  $(\varepsilon_{nt}, \varepsilon_{st})'$  is assumed to be independently distributed with mean 0 and covariances  $E\varepsilon_{nt}^2 = \sigma^2$ ,  $E\varepsilon_{nt}\varepsilon_{st} = 0$ , and  $E\varepsilon_{st}^2 = \mu\sigma^2$ . When  $\rho_{n2}$  and  $\rho_{s2}$  are assumed to be zero, ML estimation of (4.12) and (4.13), given annual data on  $n_t$  and  $s_t$  and quarterly data on  $u_t = s_t - n_t$ , gives

$$\begin{aligned} \hat{\rho}_{n1} &= .911, & \hat{\rho}_{s1} &= .898, & \hat{c}_n &= .000, \\ &(.041) & &(.019) & &(.008) \end{aligned}$$

$$\begin{aligned} \hat{c}_s &= .012, & \hat{\mu} &= .310, & \hat{\sigma}^2 &= 2.5 \times 10^{-5}, \\ &(.004) & &(.034) & & \end{aligned} \quad (4.14)$$

a log-likelihood value of 271.4,  $r_1 = .25$ ,  $r_2 = .18$ ,  $r_3 = .09$ , and  $r_4 = -.16$ , where the  $r_i$ 's are estimated residual autocorrelations for  $u_t$ ,  $t \in T_1$ . The approximate standard error for the residual autocorrelations is  $T^{-1/2} = .13$ . For the unrestricted model (4.12) and (4.13), the ML estimates are

$$\begin{aligned} \hat{\rho}_{n1} &= 1.240, & \hat{\rho}_{n2} &= -.371, & \hat{\rho}_{s1} &= 1.216, \\ &(.146) & &(.134) & &(.099) \\ \hat{\rho}_{s2} &= -.367, & \hat{c}_n &= .001, & \hat{c}_s &= .011, \\ &(.074) & &(.005) & &(.003) \\ \hat{\mu} &= .261, & \hat{\sigma}^2 &= 2.0 \times 10^{-5}, & & \\ & & & & &(.019) \end{aligned} \quad (4.15)$$

with a log-likelihood value of 277.3,  $r_1 = -.14$ ,  $r_2 = .13$ ,  $r_3 = .16$ , and  $r_4 = .19$ . The bivariate AR(2) model (4.15)



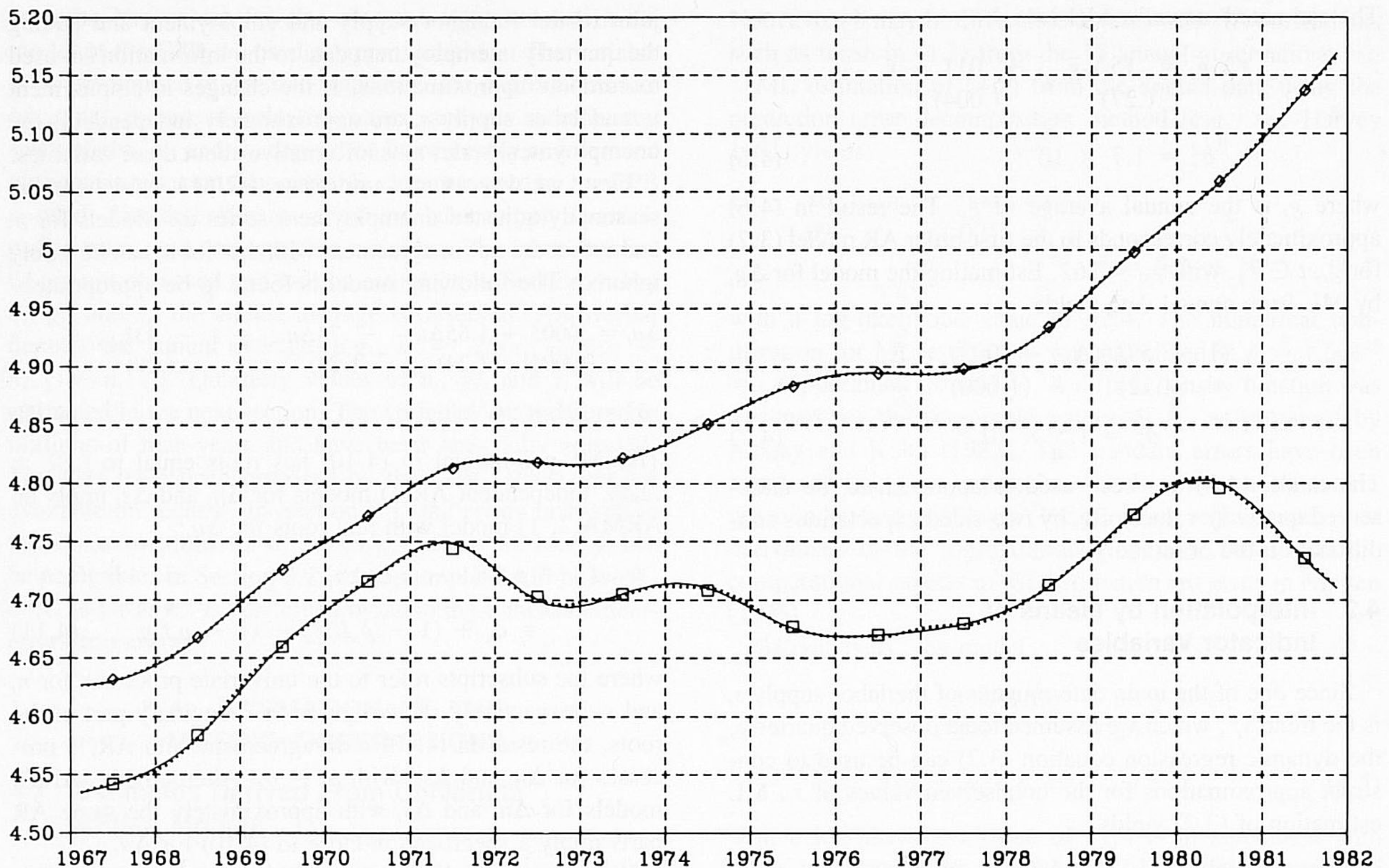


Figure 1. Total Employment and Labor Supply: Approximated Quarterly Data and Annual Observations in Millions of Man-Years. Employment annual observations ( $\square$ ): —, approx. (4.4); ···, approx. (4.15). Labor supply annual observations ( $\diamond$ ): —, approx. (4.9); ···, approx. (4.15).

differs significantly from the AR(1) model (4.14), which is not surprising, given the values of the residual autocorrelations in (4.14). We choose the bivariate model (4.15) to compute interpolations for the missing observations and their derivatives with respect to the parameters in (4.15) by means of the fixed interval smoother. The approximations for the missing observations on  $n_t$  and  $s_t$  generated by the bivariate model (4.15) and the univariate models (4.4) and (4.9) are given in Figure 1. At first sight, alternative estimates of the missing observations appear to be close to each other, which is not surprising because they are restricted to sum to the observed annual figure. The asymptotic standard errors of the approximation error of proxies generated by the univariate models are .0074 and .0030 for employment and labor supply, respectively. For the bivariate model (4.15), we get a standard error of .0027 for both series. The differences between alternative estimates of the missing observations are in some quarters statistically significant. The use of unemployment data considerably improves the accuracy of the approximations. Moreover, the serial correlation properties of these approximations and the estimates that are obtained when they are used as proxies for unobserved regressors can be quite different (see also Wilcox 1983 and Sec. 5).

An obvious extension of (4.15) consists in the inclusion of the trend in labor supply  $s_t^*$  as an indicator of  $s_t$ , as done in (4.8). Joint estimation of (3.2) and (4.12) subject to the restriction  $\gamma_{21} + \delta_2 = 1$ , given annual observations on  $n_t$

and  $s_t$  and quarterly data for  $u_t$  and  $s_t^*$ , yields

$$\begin{aligned} \hat{\rho}_{n1} &= 1.196, & \hat{\rho}_{n2} &= -.305, & \hat{c}_n &= .000, \\ & (.019) & & (.026) & & (.000) \\ \hat{\mu} &= .317, & \hat{\gamma}_{21} &= .839, & \hat{\beta}_2 &= -.025, \\ & (.024) & & (.013) & & (.004) \\ \hat{c}_2 &= .136, & \hat{\sigma}^2 &= 2.0 \times 10^{-5}, & & \\ & (.015) & & & & \end{aligned} \quad (4.16)$$

a log-likelihood value of 277.6,  $r_1 = .00$ ,  $r_2 = .14$ ,  $r_3 = .14$ , and  $r_4 = .17$ . The asymptotic standard errors of the estimation error of the missing observations generated by (4.16) are .0025 for both series. Apparently, the information in  $s_t^*$  is not important for predicting  $s_t$ .

Note, finally, that for the models we estimate, independence of the innovations in  $n_t$  and  $s_t$  has been assumed. This assumption is only approximately in accordance with the structural model presented in the previous section and could be omitted. However, no empirical results on this extension are available at present.

## 5. THE USE OF ESTIMATES OF THE MISSING DATA

### 5.1 Structural Analysis of the Adjustment Equation of Labor Demand

In this section we analyze Equation (3.1) for actual labor demand  $d_t$ . As  $d_t$  is not directly observed for any period,



(3.1) cannot be estimated by ML. If the  $a_{it}$ 's were constant within the year, the annual average of  $d_t$  could be computed from (3.3), because the annual averages of  $n_t$ ,  $g_t$ , and  $s_t$  are observed. We use (3.3) to compute proxy values for  $d_t$ :

$$d_t = [(n_t - g_t) - a_{1t}(s_t - g_t)]/a_{2t}, \quad (5.1)$$

where the weights  $a_{it}$  obtained by the CPB and the interpolations for  $n_t$ ,  $g_t$ , and  $s_t$  derived in the previous section are used. Actually, one might add a disturbance term to (5.1) to account for measurement errors in  $d_t$  due to the approximate character of (3.3). This point has not been investigated. The series  $l_t = n_t - g_t$ ,  $sb_t$ ,  $d_t$ , and  $d_t^*$  are given in Figure 2. The proxies are obtained from (4.7) and (4.15). Notice the increase of employment in 1979–1980, a period in which the desired labor demand is smaller than employment and steadily decreases. The temporary increase of actual labor demand estimated by (5.1) in that period, however, possibly explains the increase in employment. An adjustment model that distinguishes between  $d_t$  and  $d_t^*$  seems to be more in agreement with the empirical evidence than the assumption that  $d_t = d_t^*$ .

When the proxy for  $d_t$  is substituted into (3.1), one gets

$$\gamma_1(L)\Delta\hat{d}_t = \beta_1(\hat{d} - d^*)_{t-1} + \delta_1(L)\Delta d_t^* + \varepsilon_{1t} + w_t, \quad (5.2)$$

where  $w_t = [\beta_1 L - \Delta\gamma_1(L)](d - \hat{d})_t$ . The orthogonality of

the regressors and the disturbance term in (5.2) required for the consistency of OLS estimates can be violated even if the regressors in (3.1) are orthogonal to  $\varepsilon_{1t}$ , which we assume. First, the proxies can be correlated with  $w_t$ , as not all regressors are contained in the information set that generates the proxies. For instance, the information on  $g_t$  that might be contained in  $n_t$  has not been used in (4.7). Since the main determinants of the missing observations have been taken into account, however, this correlation will be neglected. A second possible cause of correlation between regressors and the error term  $\varepsilon_{1t} + w_t$  is the use of two-sided conditional expectations to generate proxies. Although, for example,  $\Delta d_{t-1}$  and  $\varepsilon_{1t}$  are assumed to be orthogonal,  $\Delta\hat{d}_{t-1}$  might depend on  $\Delta d_t$  and therefore be correlated with  $\varepsilon_{1t}$ . This problem can be avoided by using proxies that are conditional expectations, given past observations only. For most proxies, however, the dependence on future information is expected to be small, so regressors and disturbance in (5.2) can be assumed to be at least approximately orthogonal.

Two alternative specifications for (5.2) have been estimated for the sample period 1969–1981. Proxies for  $n_t$  and  $s_t$  have been generated in three different ways: (a) from the univariate models (4.4) and (4.9), (b) from the bivariate AR model (4.15), and (c) from the bivariate ARX model (4.16). In all of these cases, approximations for  $g_t$  are obtained from the AR model (4.7). The three sets of proxies and the proxies constructed by the CPB are given in Table 1.

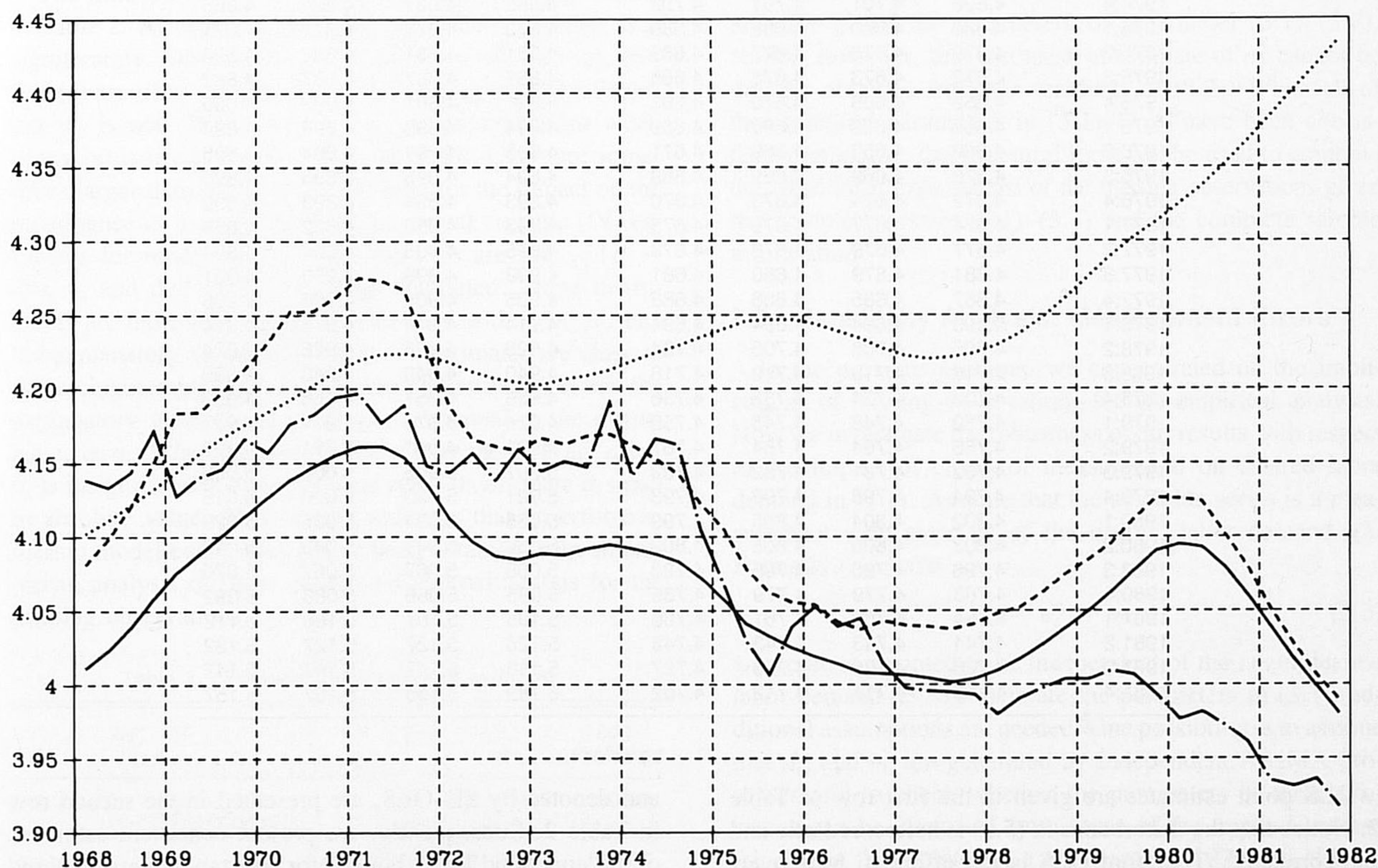


Figure 2. Labor Demand in the Private Sector and Some Related Series: Approximated Quarterly Data in Millions of Man-Years. —, Employment; ···, labor supply; ---, labor demand; — · —, desired labor demand.



Table 1. *Approximated Quarterly Data on Employment and Labor Supply in Millions of Man-Years*

Quarter	Employment				Labor Supply			
	(4.4)	(4.15)	(4.16)	CPB	(4.9)	(4.15)	(4.16)	CPB
1967:1	4.535	4.535	4.535	4.563	4.627	4.627	4.627	4.654
1967:2	4.539	4.540	4.540	4.543	4.629	4.628	4.628	4.632
1967:3	4.544	4.544	4.544	4.534	4.633	4.633	4.633	4.623
1967:4	4.551	4.549	4.549	4.528	4.639	4.640	4.640	4.619
1968:1	4.560	4.558	4.558	4.547	4.648	4.649	4.649	4.638
1968:2	4.574	4.572	4.572	4.571	4.660	4.660	4.660	4.659
1968:3	4.591	4.591	4.591	4.595	4.674	4.674	4.674	4.678
1968:4	4.611	4.615	4.615	4.624	4.690	4.689	4.689	4.699
1969:1	4.633	4.637	4.636	4.620	4.706	4.705	4.704	4.688
1969:2	4.653	4.653	4.653	4.652	4.720	4.721	4.721	4.720
1969:3	4.670	4.668	4.668	4.674	4.733	4.734	4.734	4.740
1969:4	4.685	4.683	4.683	4.692	4.745	4.745	4.745	4.755
1970:1	4.697	4.697	4.697	4.698	4.755	4.755	4.755	4.757
1970:2	4.709	4.711	4.711	4.712	4.766	4.765	4.765	4.766
1970:3	4.722	4.722	4.722	4.720	4.777	4.777	4.777	4.775
1970:4	4.735	4.734	4.734	4.732	4.790	4.790	4.790	4.788
1971:1	4.746	4.744	4.744	4.750	4.802	4.802	4.802	4.808
1971:2	4.750	4.750	4.750	4.753	4.812	4.812	4.812	4.814
1971:3	4.746	4.748	4.747	4.746	4.818	4.818	4.817	4.815
1971:4	4.734	4.735	4.734	4.729	4.821	4.821	4.820	4.815
1972:1	4.719	4.716	4.716	4.710	4.820	4.822	4.822	4.815
1972:2	4.705	4.704	4.704	4.704	4.819	4.819	4.819	4.818
1972:3	4.696	4.697	4.697	4.699	4.817	4.817	4.817	4.819
1972:4	4.693	4.695	4.696	4.699	4.816	4.815	4.816	4.820
1973:1	4.697	4.701	4.701	4.705	4.817	4.815	4.815	4.820
1973:2	4.702	4.700	4.701	4.700	4.819	4.819	4.820	4.818
1973:3	4.708	4.707	4.707	4.706	4.823	4.824	4.824	4.822
1973:4	4.713	4.712	4.712	4.710	4.829	4.830	4.830	4.828
1974:1	4.715	4.711	4.711	4.710	4.837	4.838	4.838	4.838
1974:2	4.713	4.711	4.711	4.709	4.846	4.847	4.847	4.845
1974:3	4.707	4.709	4.709	4.710	4.856	4.855	4.855	4.856
1974:4	4.698	4.701	4.701	4.702	4.865	4.864	4.864	4.865
1975:1	4.688	4.690	4.689	4.688	4.873	4.873	4.872	4.871
1975:2	4.679	4.676	4.677	4.689	4.881	4.881	4.882	4.894
1975:3	4.672	4.673	4.673	4.668	4.887	4.887	4.887	4.882
1975:4	4.669	4.669	4.670	4.663	4.891	4.891	4.892	4.885
1976:1	4.668	4.668	4.669	4.669	4.894	4.893	4.894	4.894
1976:2	4.669	4.669	4.669	4.671	4.895	4.894	4.894	4.896
1976:3	4.670	4.669	4.669	4.669	4.894	4.895	4.895	4.895
1976:4	4.672	4.674	4.673	4.670	4.893	4.894	4.893	4.890
1977:1	4.674	4.677	4.676	4.673	4.893	4.893	4.892	4.889
1977:2	4.677	4.679	4.678	4.678	4.895	4.895	4.894	4.894
1977:3	4.681	4.679	4.680	4.681	4.899	4.899	4.900	4.901
1977:4	4.687	4.685	4.686	4.688	4.905	4.905	4.906	4.908
1978:1	4.695	4.694	4.694	4.694	4.914	4.914	4.914	4.914
1978:2	4.705	4.706	4.706	4.704	4.926	4.926	4.926	4.924
1978:3	4.718	4.719	4.719	4.718	4.940	4.940	4.940	4.939
1978:4	4.733	4.734	4.733	4.736	4.956	4.957	4.956	4.959
1979:1	4.750	4.748	4.748	4.750	4.974	4.974	4.974	4.977
1979:2	4.766	4.764	4.764	4.767	4.991	4.991	4.991	4.994
1979:3	4.782	4.781	4.782	4.782	5.007	5.006	5.007	5.006
1979:4	4.794	4.798	4.798	4.793	5.021	5.020	5.020	5.015
1980:1	4.802	4.804	4.805	4.799	5.035	5.034	5.035	5.029
1980:2	4.803	4.805	4.805	4.802	5.050	5.049	5.049	5.046
1980:3	4.796	4.795	4.795	4.798	5.066	5.067	5.067	5.070
1980:4	4.783	4.779	4.779	4.785	5.085	5.086	5.086	5.092
1981:1	4.764	4.762	4.761	4.766	5.105	5.107	5.106	5.112
1981:2	4.744	4.743	4.743	4.748	5.126	5.127	5.127	5.132
1981:3	4.726	4.728	4.728	4.727	5.148	5.147	5.147	5.147
1981:4	4.710	4.711	4.712	4.702	5.169	5.166	5.167	5.157

OLS point estimates are given in the first row of Table 2. Notice that the disturbance in (5.2) is heteroscedastic and autocorrelated. Therefore OLS is not efficient. Moreover, the estimation of standard errors (SE's) of OLS estimates can be intricate (see the Appendix). "Standard errors," obtained using the standard formula for OLS standard errors

and denoted by SE-OLS, are presented in the second row in Table 2. Subsequently, we present consistent estimates of the upper and lower bounds for the standard errors based on straightforward extensions of the results in the Appendix (details can be found in Nijman and Palm 1984). We give the White-Domowitz (1984) estimates for the standard er-



Table 2. The Adjustment Equation (5.2) for Labor Demand

	Model-Based Proxies												Proxies Generated by the CPB			
	(a)				(b)				(c)							
	$\beta_1$	$\gamma_{11}$	$\gamma_{12}$	$\delta_1$	$\beta_1$	$\gamma_{11}$	$\gamma_{12}$	$\delta_1$	$\beta_1$	$\gamma_{11}$	$\gamma_{12}$	$\delta_1$	$\beta_1$	$\gamma_{11}$	$\gamma_{12}$	$\delta_1$
	$\gamma_{12} = 0$															
OLS	−.036	.789		.092	−.039	.700		.139	−.039	.705		.138	−.014	.674		.222
SE-OLS	.016	.071		.051	.019	.082		.062	.019	.081		.061	.020	.086		.067
WDO-0	.014	.079		.067	.019	.095		.074	.018	.094		.074				
WDO-4	.016	.078		.074	.018	.088		.075	.018	.087		.075				
	Unrestricted estimates															
OLS	−.036	.773	.019	.092	−.041	.652	.060	.140	−.040	.675	.037	.138	−.021	.517	.210	.223
SE-OLS	.016	.134	.133	.051	.019	.129	.127	.061	.019	.129	.127	.061	.020	.125	.124	.065
WDO-0	.015	.121	.112	.068	.019	.135	.122	.075	.018	.134	.124	.074				
WDO-4	.016	.119	.130	.074	.017	.130	.131	.074	.017	.127	.131	.074				

NOTE: OLS = ordinary least squares; SE-OLS = OLS standard error; WDO = White-Domowitz estimate for standard error.

rors (WDO- $r$ ), which only take account of nonzero elements on the main diagonal (heteroscedasticity) when  $r = 0$  and up to a four-period dependence in the disturbance covariance matrix when  $r = 4$ . Upper and lower bounds for the SE's have been computed for  $r = 4$ . Since the first four decimal points of these bounds equal WDO-4, they have not been reported separately. The SE's are found to be hardly sensitive to changes in  $r$ . More important, it appears that the effect on the SE's of estimating parameters in the proxies is almost negligible. If this holds true more generally, the computation of SE's could be greatly simplified.

The following conclusions can be drawn from the results in Table 2. As expected, the estimate of  $\beta_1$  is negative. It is significantly different from zero when the proxies (a), (b), and (c) are used. The coefficients  $\beta_1$  and  $\gamma_{11}$  are significant, but  $\gamma_{12}$  is not. The coefficient  $\delta_1$  is not significant when proxy (a) is used. The algebraic SE's (SE-OLS) are sometimes larger than the SE's that account for the impact on the disturbance of using proxies. Instrumental variable (IV) estimates for model (5.2), where the IV's are  $\Delta\hat{d}_{t-5}$ ,  $(\hat{d} - d^*)_{t-5}$ , and  $\Delta d_t^*$ , have also been computed. These instruments are more likely to be uncorrelated with  $\varepsilon_{1t} + w_t$  than the explanatory variables in (5.2). IV estimates are close to those obtained by OLS, confirming our assumption that the explanatory variables in (5.2) are orthogonal to the disturbance term. When the proxies derived by the CPB are used,  $\beta_1$  is insignificantly different from zero. Its estimate is small in absolute value. This finding indicates that an error-correction model could hardly have been obtained from an empirical analysis of these data-based approximations for the missing observations.

Table 3. Estimates of Model (5.5) and (5.6)

	$\rho$	$c$	$\phi$	$\mu$	$\sigma^2$	Log-likelihood
$\rho = \phi = 0$		-.004 (.002)		7.90 (1.83)	$3.9 \times 10^{-5}$	130.98
$\rho = 0$		-.004 (.002)	-.43 (.36)	15.51 (3.56)	$2.0 \times 10^{-5}$	131.17
$\phi = 0$	.35 (.23)	-.004 (.003)		1.58 (.76)	$1.0 \times 10^{-4}$	131.19

To validate the model (5.2), one can investigate the response of actual labor demand to a shift in desired labor demand. Error-correction models completely adjust to a step change in the target variable. When the mean lag is 0, they also completely adjust to a trend in the target. The mean lag  $(\delta_1 + \gamma_{11} + \gamma_{12} - 1)/\beta_1$  can be readily estimated from the results in Table 2. For the specifications reported in Table 2, it is much larger than 1. However, the variable  $d_t^*$  is at most locally trending in the sample period, so in this respect, the choice of an error-correction model with a nonzero mean lag may be reasonable for the period 1969–1981.

With the exception of  $\sigma_1^2$ , consistent estimates are available for all of the parameters of the model (3.1)–(3.7). Notice, however, that a consistent estimate of  $\sigma_1^2$  cannot be directly obtained from the residuals of (5.2). When all of the structural parameters in (3.1)–(3.7) have been consistently estimated, the structural form can be used to compute the conditional expectations of the missing observations given the complete system (3.1)–(3.7) and the complete sample information.

## 5.2 Sensitivity Analysis: Measurement Errors

In the previous sections, we concentrated on the implications of missing observations for the empirical analysis. Now we investigate the robustness of our results with respect to the impact of errors of measurement on desired labor demand in (3.1). Assume that the available series is a measurement with error  $d_t^m$  of the desired labor demand  $d_t^*$ . Formally, we have

$$d_t^m = d_t^* + m_t, \quad (5.3)$$

where  $m_t$  is assumed to be independent of the latent desired labor demand  $d_t^*$ . To estimate the parameters in (3.1), additional assumptions are needed. One possibility is to assume that  $d_t^*$  and  $m_t$  are generated by independent ARIMA processes and to check whether the parameters of these processes can be identified from those of the implied process for  $d_t^m$  (on this point see, e.g., Maravall 1979 and Nijman 1985).

The measurements of desired labor demand,  $d_t^m$ , are approximately generated by a random walk. Estimation of the



Table 4. The Model for  $d_t^m$ 

Restrictions	$\hat{\rho}_1$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{c}$	$\hat{\sigma}_v^2$
$\rho = \phi = 0$		-.10		-.004	$3.8 \times 10^{-4}$
$\rho = 0$	-.43	.30		-.005	$3.7 \times 10^{-4}$
$\phi = 0$	.35	-.45	-.09	-.004	$3.8 \times 10^{-4}$

slightly more general IMA(1, 1) model yields

$$\Delta d_t^m = \hat{v}_t - .10\hat{v}_{t-1} - .004, \quad (5.4)$$

(.14)      (.003)

$$\hat{\sigma}_v^2 = 4.4 \times 10^{-4}.$$

Several models for  $d_t^*$  and  $m_t$  are in accordance with the empirical evidence in (5.4). To illustrate this point, consider the models

$$(1 - \rho L)\Delta d_t^* = v_{1t} + c \quad (5.5)$$

and

$$(1 - \phi L)m_t = v_{2t}, \quad (5.6)$$

where  $v_{1t}$  and  $v_{2t}$  are independent white noise processes with variances  $\mu\sigma^2$  and  $\sigma^2$ , respectively.

ML estimation subject to the restrictions (a)  $\rho = \phi = 0$ , (b)  $\rho = 0$ , and (c)  $\phi = 0$ , respectively, yields the results reported in Table 3. The economic implications of the three models are markedly different. The ratio of the variances of  $\Delta d_t^*$  and  $m_t$ ,  $\mu(1 - \phi^2)(1 - \rho^2)^{-1}$ , equals 7.90, 12.64, and

Table 5. Estimates for the Labor Demand Equation (5.2) Subject to Errors of Measurement, Proxy Set (c)

Model for $d_t^m$	$\beta_1$	$\gamma_{11}$	$\delta_1$
$\rho = \phi = 0$			
OLS	-.036	.692	.188
SE-OLS	.019	.081	.075
WDO-0	.018	.096	.088
WDO-4	.018	.090	.088
$\rho = 0$			
OLS	-.035	.693	.194
SE-OLS	.019	.080	.072
WDO-0	.018	.096	.085
WDO-4	.017	.090	.083
$\phi = 0$			
OLS	-.028	.649	.328
SE-OLS	.019	.082	.108
WDO-0	.018	.099	.116
WDO-4	.016	.095	.117

1.80 for the three models, respectively. We write the model for  $d_t^m$  as  $(1 - \rho_1 L)\Delta d_t^m = c + (1 + \theta_1 L + \theta_2 L^2)v_t$ . The estimates implied by these three models, given in Table 4, are all close to each other and to the estimates reported in (5.4). It is obvious that the data cannot discriminate among the three models (5.5). Such a situation is expected to arise frequently in the presence of unobservables. At present, we are primarily interested in the parameters of (5.2). Therefore we investigate the sensitivity of the parameter estimates of (5.2) with respect to assumptions on  $\rho$  and  $\phi$ . Moreover,



Figure 3. Desired Labor Demand by the Private Sector: Observations and Corrections for Measurement Errors in Millions of Man-Years. —, Observations. ···, approxs. (5.5) and (5.6),  $\rho = \phi = 0$ . ---, approxs. (5.5) and (5.6),  $\rho = 0$ . — · —, approxs. (5.5) and (5.6),  $\phi = 0$ .



we compare the results with those obtained in Section 5.1, where we neglected measurement error in  $d_t^*$ . The expected values of  $d_t^*$ , given all observations on  $d_t^m$ , are displayed in Figure 3. Not surprisingly, model (c) implies the largest corrections to measured data.

The parameter estimates of (5.2) for  $\gamma_{12} = 0$  are given in Table 5. Only the set of proxies (c) described in Section 4.3 is used with a proxy for  $d_t^*$  generated by models (a), (b), and (c), respectively. The significance of coefficients is hardly affected by the way in which the measurement errors are treated. The estimate of  $\delta_1$  substantially increases when the specification (c) is used. Similar conclusions hold true for other versions of (5.2), the results for which are not reported here. Therefore we can conclude that the empirical results of Section 5.1 are fairly robust with respect to the impact of possible measurement errors in  $d_t^*$ .

## 6. CONCLUSION

In this article we showed how model-based approximations for the missing observations can be obtained and how these approximations can be used as proxies in a subsequent econometric analysis. The information set from which the approximations were derived was gradually extended. Using univariate and bivariate models, we generated proxies for the quarterly total employment, employment in the public sector, and the labor supply in the Netherlands. In the bivariate models we made use of the quarterly observations on unemployment to improve the accuracy of the proxies. This is an extension of interpolation procedures proposed in the literature.

The proxies were substituted in a structural equation for the adjustment of labor demand. The resulting equation was analyzed by methods that take account of the approximation error inherent in the proxies for the missing observations. These methods were subsequently applied to investigate the sensitivity of the empirical findings with respect to the effect of possible measurement errors in the explanatory variable of the adjustment equation for labor demand. Standard errors for the estimated regression coefficients were computed using the methods presented in the Appendix. The effect of estimated parameters in the models for the proxies on the finally reported standard errors was small in this example. It is important for applied work to investigate whether this holds true in general. Finally, the dynamic properties of the labor demand equation were analyzed. The models performed reasonably well.

## ACKNOWLEDGMENTS

An earlier version of this article was written when the authors were affiliated with the Free University, Amsterdam. Some of the work of the first author was done when he was at the Netherlands Central Bureau of Statistics. The authors thank a referee for his useful comments and R. J. Reichardt for his help in carrying out the computations. The views expressed in this article are those of the authors and do not necessarily reflect the policies of the Netherlands Central Bureau of Statistics.

## APPENDIX

For simplicity, we assume that  $y_t$  in (2.1) is observed for all periods  $t$  and is explained by one exogenous variable  $x_t$ ;  $\bar{x}_t$  is linear in  $\theta$ , say,  $\bar{x}_t = z_t' \theta$ ; and  $\hat{x}_t$  is orthogonal to  $u_t$  in large samples, so ordinary least squares are consistent for (2.3). The error term in (2.3) can be written as

$$u_t = u_{1t} + u_{2t} \quad (\text{A.1})$$

with  $u_{1t} = \varepsilon_t + (x_t - \bar{x}_t)\beta$  and  $u_{2t} = (\bar{x}_t - \hat{x}_t)\beta = \beta z_t'(\theta - \hat{\theta})$ . The large sample variance of the OLS estimator based on the proxy  $\hat{x}_t$  is given by

$$\text{var}(\sqrt{T}\hat{\beta}_p) = \text{plim}(\hat{x}'\hat{x}/T)^{-1}A(\hat{x}'\hat{x}/T)^{-1}, \quad (\text{A.2})$$

where  $\hat{x}$  is a  $T \times 1$  column vector with typical element  $\hat{x}_t$ . The matrix  $A$  is defined as  $A = A_{11} + A_{12} + A_{21} + A_{22}$ , where  $A_{ij} = \hat{x}'\Omega_{ij}\hat{x}$  and  $\Omega_{ij} = Eu_iu_j'$  (assumed to exist). The effect of estimation of  $\theta$  on the variance of  $\hat{\beta}_p$  is not asymptotically negligible and has to be taken into account.

In many applications, one can obtain  $\Omega_{11}$  as a function of the parameters of the model and substitute a consistent estimate for these parameters to get a consistent estimate of  $A_{11}$ . Alternatively, consider the White and Domowitz (1984) estimator, which reads in our notation as

$$\hat{A}_{11} = T^{-1} \left[ \sum_t \hat{u}_t^2 \hat{x}_t^2 + \sum_{\tau=1}^T \sum_t \hat{u}_t \hat{u}_{t-\tau} (\hat{x}_t \hat{x}_{t-\tau} + \hat{x}_{t-\tau} \hat{x}_t) \right] \quad (\text{A.3})$$

with  $\hat{u}_t = y_t - \hat{\beta}_p' \hat{x}_t$ . Notice that although an estimate of  $u_t$  is substituted in (A.3),  $\hat{A}_{11}$  will generally converge to  $A_{11}$  and not to  $A$ . Loosely speaking, the estimator  $\hat{A}_{11}$  does not take account of the estimated parameters in the proxy. (For more details, see Nijman and Palm 1984.)

Provided certain regularity conditions are fulfilled,  $A_{22}$  can be consistently estimated by

$$\hat{A}_{22} = T^{-2} \hat{x}' z \hat{V}_0 z' \hat{x}, \quad (\text{A.4})$$

where  $\hat{V}_0$  is a consistent estimate of the asymptotic covariance matrix of  $\sqrt{T}\hat{\theta}$ . Estimation of  $A_{12} = A_{21}'$  can cause problems. In applied work, one may be satisfied with lower and upper bounds for the asymptotic variance of  $\hat{\beta}_p$  in (A.2), which can be obtained without estimating  $A_{12}$ .

Define

$$B_{ij} = \text{plim}(\hat{x}'\hat{x}/T)^{-1}A_{ij}(\hat{x}'\hat{x}/T)^{-1}.$$

The  $B_{ij}$ 's are the large sample variances of the two components of

$$\hat{\beta}_p - \beta = (\hat{x}'\hat{x}/T)^{-1}[\hat{x}'u_1/T + \hat{x}'u_2/T]$$

and the covariance between these components. Using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} B_{11} + B_{22} - 2B_{11}^{1/2}B_{22}^{1/2} &\leq \text{var}(\sqrt{T}\hat{\beta}_p) \\ &\leq B_{11} + B_{22} + 2B_{11}^{1/2}B_{22}^{1/2}. \end{aligned} \quad (\text{A.5})$$

Consistent estimates of the bounds can be readily obtained



when consistent estimates of  $A_{11}$  and  $A_{22}$  are available. Alternative ways to compute standard errors and extensions to more general models with missing observations and other types of unobservables can be found in Nijman and Palm (1984), where results on the efficiency of various proxy variable estimates are also given.

[Received August 1984. Revised June 1985.]

## REFERENCES

- Anderson, B. D. O., and Moore, J. B. (1979), *Optimal Filtering*, New York: Prentice-Hall.
- Ansley, C. F., and Kohn, R. (1983), "On the Estimation of ARIMA-Models With Missing Values," technical report, University of Chicago, Graduate School of Business.
- Boot, J. C. G., Feibes, W., and Lisman, J. H. C. (1967), "Further Methods of Derivation of Quarterly Figures From Annual Data," *Applied Statistics*, 16, 65-75.
- Central Planning Bureau (1983), *Kompas, Kwartaalmodel Voor Prognose, Analyse en Simulatie* (Monograph 26), The Hague, The Netherlands.
- Chow, G. C., and Lin, A. (1971), "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series," *Review of Economics and Statistics*, 53, 372-375.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Ser. B*, 39, 1-38.
- Fernandez, R. B. (1981), "A Methodological Note on the Estimation of Time Series," *Review of Economics and Statistics*, 63, 471-476.
- Ginsburgh, V. A. (1973), "A Further Note on the Derivation of Quarterly Figures Consistent With Annual Data," *Applied Statistics*, 22, 368-374.
- Harvey, A. C. (1981), *Time Series Models*, Oxford: Philip Allan.
- Harvey, A. C., and Pierse, R. G. (1984), "Estimating Missing Observations in Economic Time Series," *Journal of the American Statistical Association*, 79, 125-131.
- Kloek, T. (1984), "Dynamic Adjustment When the Target Is Nonstationary," *International Economic Review*, 25, 315-326.
- Kooiman, P., and Kloek, T. (1979), "Aggregation of Micro Markets in Disequilibrium," technical report, Erasmus University Rotterdam, Econometric Institute.
- Litterman, R. B. (1983), "A Random Walk, Markov Model for the Distribution of Time Series," *Journal of Business & Economic Statistics*, 1, 169-173.
- Maravall, A. (1979), *Identification in Dynamic Shock—Error Models*, Berlin: Springer-Verlag.
- Nijman, T. E. (1985), "Missing Observations in Dynamic Macroeconomic Modeling," unpublished Ph.D. dissertation, Free University, Amsterdam, Dept. of Economics.
- Nijman, T. E., and Palm, F. C. (1984), "Consistent Estimation of Models With Unobserved Variables," Research Report 137, Free University, Amsterdam, Dept. of Econometrics.
- Pagan, A. R. (1984), "Econometric Issues in the Analysis of Regressions With Generated Regressors," *International Economic Review*, 25, 221-248.
- Palm, F. C., and Nijman, T. E. (1984), "Missing Observations in the Dynamic Regression Model," *Econometrica*, 52, 1415-1436.
- Salmon, M. (1982), "Error Correction Mechanism," *Economic Journal*, 92, 615-629.
- Watson, M. W., and Engle, R. F. (1983), "Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression Models," *Journal of Econometrics*, 23, 385-400.
- White, H., and Domowitz, I. (1984), "Nonlinear Regression With Dependent Observations," *Econometrica*, 50, 485-500.
- Wilcox, J. A. (1983), "Disaggregating Data Using Related Series," *Journal of Business & Economic Statistics*, 1, 187-191.
- Zellner, A., and Palm, F. (1974), "Time Series Analysis and Simultaneous Equation Econometric Models," *Journal of Econometrics*, 2, 17-54.